

## BIG DATA TEXNOLOGIYALARI: HADOOP VA SPARK

Sulaymonova Dilyora Sherzodjon qizi

FarDU Axborot tizimlari va texnologiyalari yo‘nalishi 3-kurs talabasi

dilyorasulaymonova36@gmail.com

Sobirjonov Behzod Qahramonovich

FarDU Axborot texnologiyalari kafedrası Katta o‘qituvchisi

bekzodbekqahromonovich@gmail.com

<https://doi.org/10.5281/zenodo.20354989>

**Annotatsiya:** Ushbu maqolada katta hajmli ma’lumotlar (Big Data) bazalarini boshqarish uchun qo‘llaniladigan zamonaviy texnologiyalar, ularning tuzilishi, asosiy xususiyatlari va amaliy qo‘llanilish sohalari batafsil yoritilgan. Xususan, Hadoop, Spark va NoSQL kabi keng tarqalgan tizimlarning ishlash prinsiplari, vazifalari va afzalliklari chuqur tahlil qilingan. Hadoop texnologiyasining taqsimlangan saqlash va parallel qayta ishlash imkoniyatlari, Spark’ning yuqori tezlikda hisoblash va real vaqt rejimida tahlil qilish xususiyatlari, shuningdek NoSQL ma’lumotlar bazalarining moslashuvchanligi va masshtablanuvchanligi alohida ko‘rib chiqilgan. Shuningdek, Big Data bilan ishlash jarayonida yuzaga keladigan muammolar, jumladan ma’lumotlar xavfsizligi, maxfiylik, katta hajmdagi ma’lumotlarni saqlash va qayta ishlashdagi murakkabliklar hamda malakali mutaxassislar yetishmasligi kabi masalalar tahlil etilgan. Shu bilan birga, Big Data texnologiyalarining biznes, sog‘liqni saqlash, ta’lim, moliya va sanoat sohalarida samaradorlikni oshirish, to‘g‘ri qarorlar qabul qilish va raqobatbardoshlikni kuchaytirishdagi imkoniyatlari ochib berilgan. Tadqiqot natijalari Big Data texnologiyalari zamonaviy tashkilotlar uchun strategik ahamiyatga ega bo‘lgan, samaradorlikni sezilarli darajada oshiruvchi muhim vosita ekanini ko‘rsatadi.[1:23.]

**Kalit so‘zlar:** Big Data atamasi, Big Data serverlari, ma’lumotlarni tahlil qilish texnologiyasi, moliyaviy sektor, vizualizatsiya, Hadoop va Spark [1:23.]

## BIG DATA TECHNOLOGIES: HADOOP AND SPARK

Sulaymonova Dilyora Sherzodjon qizi

3rd-year student, Information Systems and Technologies Department, FarDU

dilyorasulaymonova36@gmail.com

Sobirjonov Behzod Qahramonovich

Senior Lecturer, Department of Information Technologies, FarDU

bekzodbekqahromonovich@gmail.com

**Abstract:** This article provides a detailed overview of modern technologies used for managing large-scale Big Data databases, their structure, key characteristics, and practical application areas. In particular, the operating principles, functions, and advantages of widely used systems such as Hadoop, Spark, and NoSQL are thoroughly analyzed. The distributed storage and parallel processing capabilities of Hadoop technology, Spark’s high-speed computing and real-time analytics features, as well as the flexibility and scalability of NoSQL databases are examined separately. In addition, the article analyzes the challenges encountered when working with Big Data, including data security, privacy, the complexity of storing and processing massive volumes of information, and the shortage of qualified specialists. At the same time, the possibilities of Big Data technologies in increasing efficiency, supporting accurate decision-making, and enhancing competitiveness in business, healthcare, education, finance, and industrial sectors are highlighted. The research results

demonstrate that Big Data technologies are strategically important tools for modern organizations, significantly improving operational efficiency. [1:23.]

**Keywords:** Big Data concept, Big Data servers, data analysis technology, financial sector, visualization, Hadoop and Spark. [1:23.]

## ТЕХНОЛОГИИ БОЛЬШИХ ДАННЫХ: HADOOP И SPARK

Сулаймонова Дилёра Шерзоджон кизи

Студентка 3 курса направления «Информационные системы технологий»,  
ФерГУ

dilyorasulaymonova36@gmail.com

Собиржонов Бехзод Кахрамонович

Старший преподаватель кафедры информационных технологий ФерГУ  
bekzodbekqahromonovich@gmail.com

**Аннотация.** В данной статье подробно рассматриваются современные технологии, применяемые для управления базами данных Big Data, их структура, основные характеристики и сферы практического применения. В частности, глубоко проанализированы принципы работы, функции и преимущества широко распространённых систем, таких как Hadoop, Spark и NoSQL. Отдельно рассмотрены возможности распределённого хранения и параллельной обработки данных в технологии Hadoop, особенности высокоскоростных вычислений и анализа в режиме реального времени в Spark, а также гибкость и масштабируемость баз данных NoSQL. Кроме того, проанализированы проблемы, возникающие при работе с Big Data, включая безопасность данных, конфиденциальность, сложности хранения и обработки больших объёмов информации, а также нехватку квалифицированных специалистов. Вместе с тем раскрыты возможности технологий Big Data в повышении эффективности, принятии правильных решений и усилении конкурентоспособности в сферах бизнеса, здравоохранения, образования, финансов и промышленности. Результаты исследования показывают, что технологии Big Data являются стратегически важным инструментом для современных организаций, значительно повышающим эффективность деятельности. [1:23.]

**Ключевые слова:** Концепция Big Data, серверы Big Data, технология анализа данных, финансовый сектор, визуализация, Hadoop и Spark. [1:23.]

**Kirish:** Bugungi kunda korxonalar uchun juda muhim bo'lgan Big Data ma'lumotlar to'plamiga tegishli bo'lib, ularning hajmi, tezligi va xilma-xilligi tufayli an'anaviy usullar yordamida qayta ishlanmaydi. Ushbu blog posti Big Data nima ekanligini va nima uchun muhimligini tushuntiradi, shuningdek, Hadoop va Spark kabi mashhur qayta ishlash vositalarini batafsil ko'rib chiqadi. Unda Hadoop'ning afzalliklari va kamchiliklari, Spark bilan ma'lumotlarni qayta ishlash jarayonlari va zamonaviy muqobil variantlari solishtiriladi. Shuningdek, u vositani tanlashda e'tiborga molik jihatlari, Hadoop va Spark o'rtasidagi farqlar, muvaffaqiyatli strategiyalar, ularning biznes olamiga ta'siri va samaradorlikni oshiradigan vositalarni muhokama qiladi. Oxir oqibat, to'g'ri vositalarni tanlash va Big Data loyihalari uchun samarali strategiyalarni ishlab chiqish korxonalar uchun raqobatdosh ustunlikka erishish uchun juda muhimdir. Big Data (Katta ma'lumotlar) an'anaviy ma'lumotlarni qayta ishlash dasturlari tomonidan qayta ishlanishi uchun

juda katta, murakkab va tez oqimga ega bo'lgan ma'lumotlar to'plamlarini anglatadi. Ushbu ma'lumotlar tuzilgan (masalan, ma'lumotlar bazasidagi jadvallar), tuzilmagan (matnli hujjatlar, tasvirlar, videolar) va yarim tizimli (XML, JSON fayllari) formatlarda bo'lishi mumkin. Katta ma'lumotlarning kattaligi, xilma-xilligi, tezligi va haqiqiyli (4V qoidasi) an'anaviy usullar yordamida tahlil qilishni qiyinlashtiradi. Biroq, to'g'ri vositalar va texnikalar bilan tahlil qilinganda, u korxonalariga qimmatli tushunchalar berishi va raqobatdosh ustunlikni ta'minlashi mumkin. Katta ma'lumotlarni qayta ishlash va tahlil qilish maxsus vositalar va texnologiyalarni talab qiladi. Hadoop, Spark, NoSQL ma'lumotlar bazalari va bulutga asoslangan yechimlar katta ma'lumotlarni qayta ishlash infratuzilmasining asosini tashkil qiladi. Ushbu vositalar katta ma'lumotlar to'plamlarini parallel qayta ishlash va tahlil qilish imkonini beradi, bu esa korxonalariga tez va samarali qarorlar qabul qilishda yordam beradi. Bundan tashqari, mashinani o'rganish va sun'iy intellekt algoritmlari katta ma'lumotlardagi murakkab munosabatlarni ochish va bashorat qilish uchun ishlatiladi. [1:23.]

### **Natija va muhokama**

“Katta ma'lumotlar” ning ahamiyati bugungi kunda korxonalarining qaror qabul qilish jarayonlarini yaxshilashidan kelib chiqadi. Katta ma'lumotlarni tahlil qilish ko'plab sohalarda, jumladan, mijozlar xatti-harakatlarini yaxshiroq tushunish, marketing strategiyalarini optimallashtirish, operatsion samaradorlikni oshirish va xavflarni kamaytirishda qo'llanilishi mumkin. Misol uchun, chakana savdo kompaniyasi qaysi mahsulotlar birgalikda sotilishini aniqlash uchun mijozlarning xarid qilish odatlarini tahlil qilishi va shunga mos ravishda do'kon tartibini optimallashtirishi mumkin. Xuddi shunday, moliyaviy institut katta ma'lumotlarni tahlil qilish orqali firibgarlik faoliyatini tezroq aniqlashi mumkin. Hadoop, Katta ma'lumotlar Bu klasterlarni qayta ishlash uchun mo'ljallangan ochiq manbali ramka. U katta hajmdagi ma'lumotlarni taqsimlangan holda saqlash va qayta ishlash uchun ishlatiladi. Apache Hadoop loyihasi ma'lumotlar bo'yicha olimlar va muhandislarga murakkab ma'lumotlarni tahlil qilish imkonini beruvchi kengaytiriladigan, ishonchli va tejamkor yechimni taqdim etadi. Hadoopning asosiy maqsadi ma'lumotlarni kichik bo'laklarga bo'lish, ularni bir nechta kompyuterlar bo'ylab tarqatish va ularni parallel ravishda qayta ishlash, natijada tezroq natijalarga erishishdir.

Hadoop-ning mashhurligi, iqtisodiy samaradorlik va Masshtablilik Bu Hadoop ekotizimlari bilan chambarchas bog'liq. Uning tovar uskunasi ishlar qobiliyati kompaniyalarga qimmatbaho ixtisoslashtirilgan apparatlarga sarmoya kiritmasdan katta ma'lumotlar loyihalarini amalga oshirish imkonini beradi. Bundan tashqari, Hadoop ekotizimlari doimiy ravishda rivojlanib, yangi vositalar va texnologiyalar bilan integratsiyalashib, Hadoopni katta ma'lumotlarni qayta ishlash maydonida asosiy o'yinchiga aylantirmoqda.

Xususiyat	Tushuntirish	Foyda
Tarqalgan ishlov berish	Ma'lumotlar bir nechta tugunlarda parallel ravishda qayta ishlanadi.	Tez va kengaytiriladigan ma'lumotlarni qayta ishlash.
HDFS (Hadoop tarqatilgan fayl tizimi)	Ma'lumotlarni taqsimlangan holda saqlaydi.	Yuqori xatoga chidamlilik va ma'lumotlarning ortiqcha bo'lishi.
MapReduce	Ma'lumotlarni qayta ishlash modeli.	Parallel ishlov berish imkoniyatlari.
YARN (Yana bir resurs muzokarachisi)	Resurslarni boshqarish va ishni rejalashtirish.	Resurslardan samarali foydalanish.

Apache Spark katta ma'lumotlarni qayta ishlash sohasida katta ma'lumotlar Spark - bu klasterlarda tez va samarali tahlil qilish imkonini beruvchi ochiq manbali tizim. Uning Hadoopning MapReduce modeliga qaraganda ancha tezroq ishlov berish tezligini amalga oshirish qobiliyati Spark-ni ma'lumotlar olimlari va muhandislari uchun ajralmas vositaga aylantirdi. Uning xotirada ishlov berish imkoniyatlari turli xil foydalanish holatlarida, jumladan, iterativ algoritmlar va real vaqtda ma'lumotlar oqimida yuqori ishlashni ta'minlaydi. Spark shunchaki ma'lumotlarni qayta ishlash mexanizmi emas, balki boy ekotizimni taklif etadi. Ushbu ekotizim SQL so'rovlari uchun Spark SQL, mashinani o'rganish uchun MLlib, grafiklarni qayta ishlash uchun GraphX va real vaqtda ma'lumotlar oqimini qayta ishlash uchun Spark Streaming kabi komponentlarni o'z ichiga oladi. Ushbu komponentlar Sparkni ko'p qirrali qiladi katta ma'lumotlar platforma va unga turli ehtiyojlar uchun yechimlar taklif qilish imkonini beradi.

Spark va Hadoop, katta ma'lumotlar Ushbu ikki texnologiya ko'pincha qayta ishlash maydonida taqqoslanadi. Hadoop katta hajmdagi fayllarni taqsimlangan holda saqlash va qayta ishlash uchun mo'ljallangan, Spark esa ko'proq ma'lumotlarni tezkor qayta ishlash va tahlil qilishga qaratilgan. Hadoopning asosiy komponenti HDFS (Hadoop Distributed File System) ma'lumotlarni ishonchli saqlaydi, Spark esa ushbu ma'lumotlarga kirish va tahlilni amalga oshiradi. Ikki texnologiyadan birgalikda foydalanish ma'lumotlarni saqlash va tezkor qayta ishlash ehtiyojlarini qondirishi mumkin.

**Xulosa:** Katta ma'lumotlar Qayta ishlash vositalari bugungi ish dunyosining ajralmas qismiga aylandi. Hadoop va Spark kabi zamonaviy texnologiyalarning paydo bo'lishi bilan ma'lumotlarni qayta ishlash jarayonlari yanada tez va samaraliroq bo'ldi. Ushbu vositalar korxonalariga katta hajmdagi ma'lumotlarni tahlil qilish, mazmunli tushunchaga ega bo'lish, yaxshiroq qarorlar qabul qilish va raqobatdosh ustunlikka ega bo'lish imkonini beradi. Kelajakda, sun'iy intellekt va mashinani o'rganish texnologiyalari integratsiyasi bilan katta ma'lumotlarni qayta ishlash vositalari yanada rivojlangan va murakkabroq muammolarni hal qilishga qodir bo'lishi kutilmoqda. Hadoop ma'lumotlarni taqsimlangan holda saqlash va qayta ishlash uchun MapReduce algoritmidan foydalanadi. Diskga asoslangan tizim bo'lib, u katta ma'lumotlar to'plamlari uchun ideal, lekin real vaqtda qayta ishlash uchun sekinroq. Boshqa tomondan, Spark xotirada ishlov berishni qo'llab-quvvatlaydi, bu uni Hadoop-dan sezilarli darajada tezroq va real vaqtda tahlil qilish uchun mos qiladi. Hadoop, birinchi navbatda, katta hajmdagi ma'lumotlarni saqlash va ommaviy qayta ishlash uchun ishlatiladi, Spark esa tezroq, ko'proq interaktiv tahlil qilish uchun afzaldir.

### Adabiyotlar, References, Литературы:

1. Mamasidiqova, I., Husanova, O., Madaminova, A., & Tojimatov, I. (2023). DATA MINING TEXNALOGIYALARI METODLARI VA BOSQICHLARI HAMDA DATA SCIENCE JARAYONLAR. *Центральноазиатский журнал образования и инноваций*, 2(3 Part 2), 18-21 [1:23.]
2. Nurmamatovich, T. I. (2021). RAQAMLI IQTISODIYOTNING GLOBALASHUV JARAYONIDA IQTISOD TARMOQLARIDA QO'LLANILISHINING ASOSIY YO'NALISHLARI. [1:23.]
3. Usmonov, B., & Rakhimov, Q. (2019). Vibration analysis of airfoil on hereditary deformable suspensions. In *E3S Web of Conferences* (Vol. 97, p. 06006). EDP Sciences.
4. Saidov, J. Moliyaviy xizmatlarda Data Science imkoniyatlari. Toshkent: O'zbekiston milliy universiteti nashriyoti, 2023.
5. Rustamov, H. Biznes qaror qabul qilishda ma'lumotlar tahlili. Farg'ona: Farg'ona davlat universiteti nashriyoti, 2019.

6. Mualliflar guruhi. “APPLICATION OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES IN THE PUBLIC CIVIL SERVICE” xalqaro ilmiyamaliy konferensiya materiallari. 2024. Sahifa 287.
7. WHO. (2021). Global strategy on digital health 2020-2025. World Health Organization.
8. Aguirre, R.R., Suarez, O., et al. (2019). Telemedicine and diabetic retinopathy: review of published screening programs. Journal of Endocrinology and Metabolism. [1:23.]
9. Azimov, A. Data Science asoslari va biznesda qo'llanilishi. Toshkent: Fan va texnologiya, 2021.